R FUTURE DIRECTIONS

BRIXEN
June 2014

Speaker

Jan Vitek

Professor of Computer Science

Purdue University, USA

PHD: University of Geneva, CH

Speciality: Programming languages, Compilers

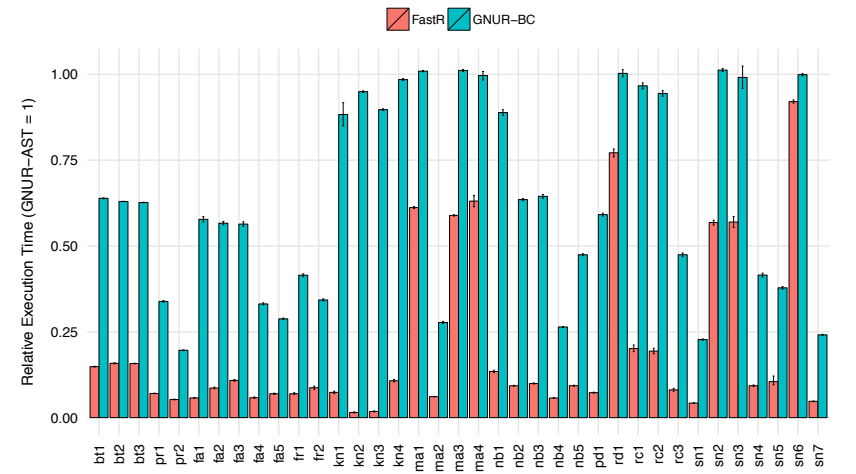
What have we done so far…

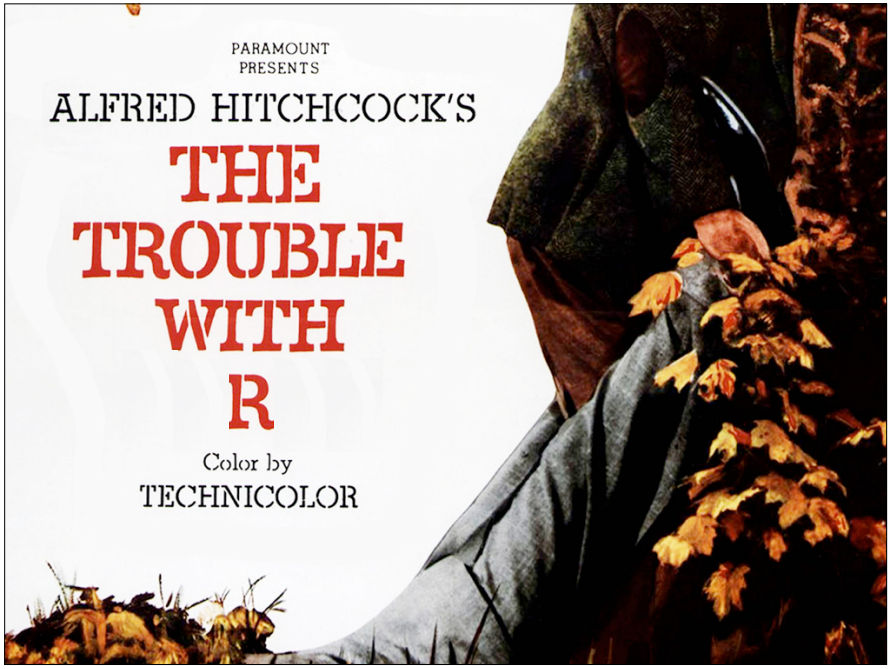
- **`TracR`** — a trace analysis / profiler for GNU-R
- **`CoreR`** — a formal semantics for a fragment of R
- **`TestR`** — a testing framework for the R language
- **`FastR`** — a new R virtual machine written in Java

Morandat, Hill, Osvald, Vitek. Evaluating the Design of the R Language. **ECOOP'12**

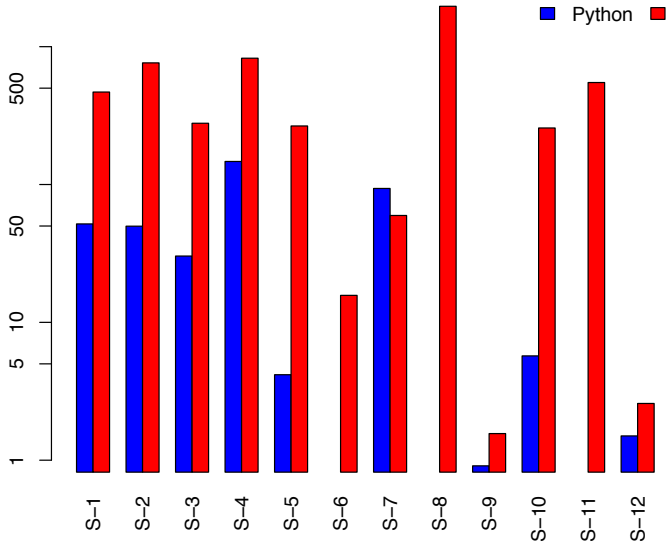
Kalibera, Maj, Morandat, Vitek. A Fast abstract syntax tree interpreter for R. **VEE'14**

FastR (speedup 8.5x GNURast)

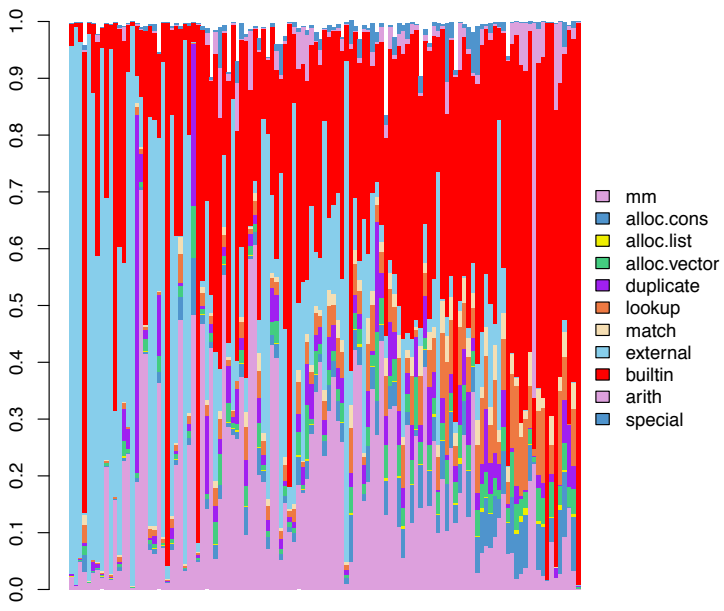




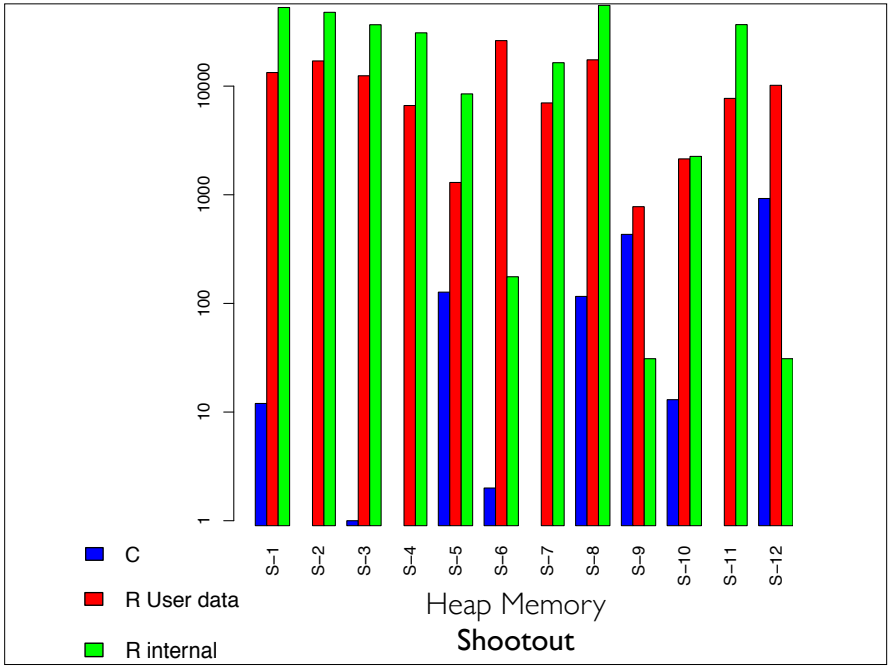
PERFORMANCE



Shootout



Time breakdown



SPECIFICATION

$$\frac{\mathsf{cpy}(H,\nu) = H',\nu' \qquad \Gamma = \iota * \Gamma' \qquad H(\iota) = F \qquad F' = F[\mathtt{x}/\nu'] \qquad H'' = H'[\iota/F']}{\mathtt{x} \mathtt{<-} \nu\, \Gamma; H \;\rightarrow\; \nu; H''}$$

45% of assignments are definitions

only 2 out of 217 million assignments are definitions in a parent frame

99.9% of side effects are local

```r
c <- 42

c(1,2,3)
```

less than 0.05% context sensitive function name lookups

only symbols that rely on it are **`c`** and **`file`**

OBJECTS

		Bioc	Misc	CRAN	Base	Total
S3	# classes	1 535	0	3 351	191	3 860
	# methods	1 008	0	1 924	289	2 438
	Avg. redef.	6.23	0	7.26	4.25	9.75
	Method calls	13M	58M	-	-	76M
	Super calls	697K	1.2M	-	-	2M
S4	# classes	1 915	2	1 406	63	2 893
	# singleton	608	2	370	28	884
	# leaves	819	0	621	16	1 234
	Hier. depth	9	1	8	4	9
	Direct supers	1.09	0	1.13	0.83	1.07
	# methods	4 136	22	2 151	24	5 557
	Avg. redef.	3	1	3.9	2.96	3.26
	Redef. depth	1.12	1	1.21	1.08	1.14
	# new	668K	64	-	-	668K
	Method calls	15M	266	-	-	15M
	Super calls	94K	0	-	-	94K

TYPES

```
x <- y

x : like D[20,20] <- y

x : D[20,20] <- y

x : D[ dim(y) ]  <- y
```

DATA LAYOUT

```
1:1000000

x[[ 300035 ]]
```

CONCURRENCY

```
x <<- 1  ||  x <<- 3
```

IMPLEMENTATIONS

	Implem. language	Project start/end	KLOC	Effort Man/ month	Percent Complete	Missing parts	Challenges	Time sinks
GNUR	C (Fortran)	1993 - …	500+	?	99.9	?	?	?
RIPOSTE	C++	2010 - …	30	40	75 Lang 10 Libs	Std libs, pkgs, search paths, locked envs, FFIs	Specification	Too large internal surface, src/main too big, messy headers, Fortran
RENJIN	Java	2010 - …	?	8	78 primitives	S4 exports, R compatibility, native code	JIT, data layout, C->R internals	
TERR	C++	2009 - …	500	264	75	R compatibility, Graphics	Clean-room imp Specification Native code	S3,S4, data.frame, model.frame, native interface
ORBIT	C (Fortran)	2011 - 2014	10	12	90+	Some bugs	Complexity of runtime	Manual specializations
FastR .168	Java	2012 - 2013	72	20	50 - 1/∞	R compatibility, graphics, internals, native interface, …	Specification	Internals, manual specialization, Truffle, benchmarks
FastR	Java	2013 - …	51	30	70 Lang 49 Libs	R compatibility, Graphics, promises, frames/envs	Specification	R compatibility, vector ops, frames/envs.


the road ahead
ROUTE
US
66